

NETWORK-BASED ANALYSIS OF A SMALL EBOLA OUTBREAK

MARK G. BURCH

College of Public Health
The Ohio State University
Columbus, OH 43210, USA

KARLY A. JACOBSEN

Mathematical Biosciences Institute
The Ohio State University
Columbus, OH 43210, USA

JOSEPH H. TIEN

Department of Mathematics and Mathematical Biosciences Institute
The Ohio State University
Columbus, OH 43210, USA

GRZEGORZ A. REMPALA

Mathematical Biosciences Institute and College of Public Health
The Ohio State University
Columbus, OH 43210, USA

ABSTRACT. We present a method for estimating epidemic parameters in network-based stochastic epidemic models when the total number of infections is assumed to be small. We illustrate the method by reanalyzing the data from the 2014 Democratic Republic of the Congo (DRC) Ebola outbreak described in Maganga et al. (2014).

1. Introduction. The best known models for the spread of infectious disease in human populations are based on the classical SIR model of Kermack and McKendrick [19]. The same system of ordinary differential equations (ODEs) may be derived as the large population limit of a density-dependent Markov jump process using the methods of Kurtz [20]. This stochastic formulation brings a number of mathematically attractive properties such as explicit likelihood formulas and ease of simulation. Nevertheless, a drawback of the Kermack and McKendrick-type models is that they can be unrealistic in describing the interactions of infectives and susceptibles as they are based on assumptions of homogeneous mixing [16].

In recent years there has been considerable interest in developing alternatives to the classical SIR, for instance, via network-based epidemic models, as reviewed by Pellis et al. [27] and House and Keeling [13]. Pair approximation models have

2010 *Mathematics Subject Classification.* Primary: 92B; Secondary: 92C60.

Key words and phrases. Ebola, Network Epidemic Models, Configuration Model, Branching Process, Statistical Inference.

This research has been supported in part by the Mathematical Biosciences Institute and the National Science Foundation under grants RAPID DMS-1513489 and DMS-0931642.

been formulated to account for spatial correlations while maintaining mathematical tractability [28, 17, 18]. Much of the work on these models has regarded the dynamics of the deterministic system obtained in the large population limit, such as the work of Miller and Volz on edge-based models [31, 24, 25] and that of Altmann [2] on a process with dynamic partnerships.

As in the work of Miller and Volz [31, 24, 25], the configuration model (CM) random graph is often chosen to dictate the network structure in epidemic models [4, 23, 5]. CM networks can be viewed as a generalization of the Erdős-Rényi random graph, i.e. the $G(n, p)$ model where n is the size of the graph and edges are drawn between each pair of nodes independently, with probability p . In the limit of a large graph this results in a Poisson degree distribution; however, in a general CM graph model framework Poisson may be replaced with any appropriate degree distribution (see, e.g., [31]). Miller and Volz heuristically formulated the limiting system of ODEs that govern the dynamics of the SIR epidemic on a CM graph. Decreusefond et al. [10] and Janson et al. [14] have recently given formal proofs for the correctness of the Miller-Volz equations as the law of large numbers (LLN) for the stochastic system under relatively mild regularity conditions on the degree distribution and on the epidemic initial condition. In addition, Volz and Miller have conjectured the limiting equations for generalizations which include dynamic graphs (edge formation and breakage), heterogeneous susceptibility, and multiple types of transmission [25]. To date, the convergence for the stochastic systems in these cases has not been formally verified but numerical studies done by the authors suggest that they are correct.

As in the case of classical SIR [1] the early behavior of an epidemic on a CM graph can be approximated by a suitable branching process [4, 14, 6]. Such approximation allows one in turn to use the early epidemic data to statistically ascertain the probability of a major outbreak (large number of infections among the network nodes) as well as to estimate the rate of infection spread and changes in the contact network as described below. By and large, there have been limited studies of statistical estimation for network-based models [32]. In one of the early papers on the topic, O'Neill and Britton study Bayesian estimation procedures for the $G(n, p)$ model when the network is assumed to be of reasonable enough size that the network structure may be included as missing data and imputed via a Monte Carlo scheme [7]. More recently, Groendyke has extended this approach to non-Markovian dynamics by allowing both the infection time and the recovery time to follow an arbitrary gamma distribution [11, 12]. However, these methods are generally tailored to networks of small size, where posterior sampling is feasible, and may encounter various difficulties in large networks when the imputation becomes computationally expensive and often impractical. The framework presented here, on the other hand, assumes the proportion of infectives is small relative to the population size. This allows us to avoid explicit imputation of the network and makes the numerical complexity of the analysis comparable to that of a small homogenous SIR epidemic [8, 29].

The main contribution of the current paper is to present a statistical inference method for analyzing the early stages of an epidemic, or a small outbreak, evolving according to SIR type dynamics on a random graph. A novel aspect of our method is that we assume the random graph structure evolves in response to the epidemic progression. This allows us to account for changing contact patterns in response to

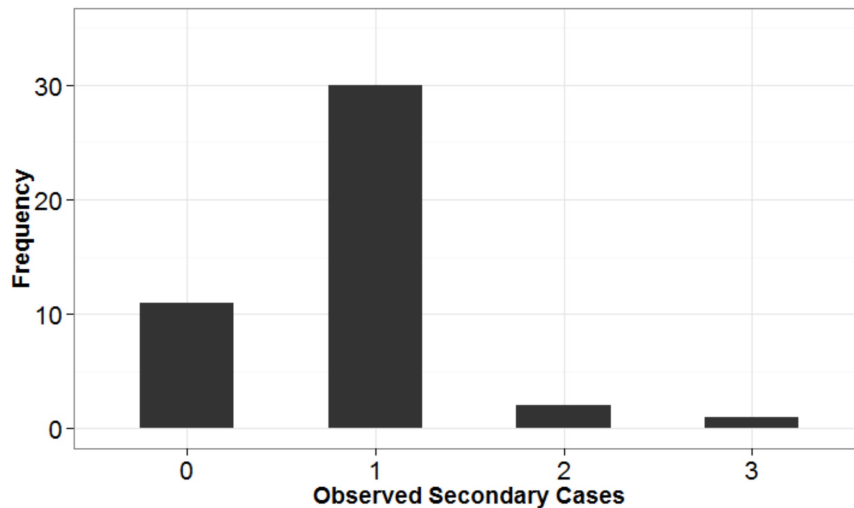


FIGURE 1. The empirical secondary case distribution in the DRC outbreak dataset (neglecting the index case), as given by Maganga et al. [22].

infection, for example due to population behavioral changes or health interventions (e.g. quarantine).

The development of our methods is motivated by the devastating 2013 – 2015 outbreak of Ebola virus in West Africa. While the West Africa outbreak received considerable attention in fall 2014 due to the dramatic rise in the number of cases, there was an independent Ebola outbreak in the Democratic Republic of the Congo (DRC) occurring at the same time. This much smaller outbreak began July 26, 2014 in Équateur Province and lasted until November 2014. The index case was a woman living in Inkanamongo village that presumably became infected by consuming bushmeat of an infected animal. Several healthcare workers involved in a postmortem cesarean section on the woman subsequently became ill and generated further chains of transmission [22].

2. DRC dataset. There were a total of 69 confirmed infections in the 2014 DRC outbreak. The time series of cumulative case counts was reported by Maganga et al. [22]. However, the analysis method presented here does not require temporal data and instead utilizes the distribution of secondary cases, also given in [22]. The index case is believed to have caused 21 secondary cases, presumably due to her funeral acting as a super-spreading event [22]. Hence, we assume this data point to be an outlier and exclude it from our final analysis, as Maganga et al. also did in their estimation of \mathcal{R}_0 . The secondary case distribution for other named contacts is shown in Figure 1. One patient caused three subsequent cases, two patients caused two additional cases, 30 caused one additional case and 11 patients caused zero additional cases. These may be viewed as observations from the post-index offspring distribution of the branching process approximation, i.e. the number of infections caused by an individual who is himself not the index case.

Although, as already mentioned, the temporal data measurements are not explicitly required for parameter estimation in our approach, the available temporal information can be directly incorporated into the likelihood function, or used to inform the parameterization of the prior distributions, which is what was done for the current DRC dataset. Details are given below.

3. Epidemic model. Consider a graph $G = \langle V, E \rangle$ where the vertex set, V , corresponds to individuals in a population of size n and the edge set, E , corresponds to potentially infectious contacts occurring between such individuals. The graph structure is given as a realization of a configuration model random graph with a prescribed degree distribution, D , where the probability that a node has degree k is denoted by $q_k \equiv P(D = k)$. Each node $i \in V$ is assigned d_i half-edges that are drawn at random from D . Then, the pool of half-edges is paired off uniformly to form the final network. This link formation model does not exclude the possibility of self-loops or multiple edges but this has a negligible effect in the large graph limit (see, e.g., discussion in [14]).

The general disease framework adopted is the standard compartmental SIR model where, for every time $t > 0$, each node $i \in V$ is classified as susceptible (S), infectious (I), or recovered (R) from the infection. Given some number of initially infectious nodes, a node i becomes infective via transmission along an edge from one of his infectious neighbors. In the Markovian case, we assume that i remains infectious for an exponentially distributed amount of time with rate parameter γ , which we refer to as the recovery rate. More generally, a non-exponential (e.g., gamma) recovery rate leads to the so-called semi-Markov SIR model [15]. While infectious, the infective i attempts to transmit the infection to all of his susceptible neighbors according to an exponential distribution with rate β . In the case that the realization of the infection “clock” for a particular neighbor “rings” before the recovery “clock”, then this neighbor becomes infected. The epidemic ends when there are no more infectives.

In addition to these standard SIR dynamics, we allow the network structure to change due to infection status. Specifically, we assume that infectious individuals drop each of their contacts according to an exponential distribution with rate δ . The dropped contacts, which could account for behavioral changes due to disease such as isolation or decreased mobility, cannot be reformed.

4. Statistical inference.

4.1. Index case offspring distribution. For an index case i with degree d_i , at most d_i secondary infections can be produced. Conditional on the recovery time of that index case, t_i , the probability that infection has passed to any particular neighbor is given by

$$p_{t_i} \equiv p(t_i; \beta, \delta) = \frac{\beta}{\beta + \delta} (1 - e^{-(\beta + \delta)t_i}) \quad (1)$$

where the first term in the product on the right-hand side is the probability that an edge transmitted infection prior to being dropped, while the second term represents the probability that either an infection or a drop occurred before recovery. Therefore, the total number of secondary infections caused by node i , which we denote

by X_i , conditional on the time of recovery t_i , is given by

$$P(X_i = x_i | t_i, d_i) = \binom{d_i}{x_i} p_{t_i}^{x_i} (1 - p_{t_i})^{d_i - x_i}.$$

If the recovery time is not known but assumed to follow a distribution $f(t_i) \equiv P(T_i = t_i)$, then we may analogously define $\pi_{d_i}(x_i)$, the conditional probability of x_i offspring given degree d_i . The law of total probability implies

$$\pi_{d_i}(x_i) = \int_0^\infty \binom{d_i}{x_i} p(t_i)^{x_i} (1 - p(t_i))^{d_i - x_i} f(t_i) dt_i.$$

We will assume identical recovery distributions for all individuals and, thus, the offspring distribution will be the same for all index cases.

The final form for the offspring distribution of an index case may be found by supposing that the degree, d_i , of the index case is unknown. Let $q_{d_i} \equiv P(D_i = d_i)$ denote the probability that the index case has degree d_i . Therefore, the law of total probability gives

$$P(X_i = x_i) = \sum_{d_i \geq x_i} q_{d_i} \pi_{d_i}(x_i). \quad (2)$$

That is, we sum over all possible degrees that could yield at least x_i secondary infections and weight them according to the degree distribution.

4.2. Post-index case offspring distribution. We now consider the offspring distribution for a post-index case. By definition, such an individual acquired infection from another individual in the network and, thus, has at least one neighbor. Therefore, some adjustments are needed to account for the fact that post-index cases have a degree distribution which differs from D . Let q'_k denote the probability that a given neighbor in the CM network has degree k . Then it is known [26] that

$$q'_k = \frac{k q_k}{\mu}.$$

Since at least one of the neighbors of a post-index case has already been infected, he may pass the infection to at most $k - 1$ of his neighbors. Let X'_i denote the offspring distribution for a post-index infection i . Similarly to Eq. (2) we derive

$$P(X'_i = x'_i) = \sum_{k > x'_i}^\infty q'_k \pi_{k-1}(x'_i). \quad (3)$$

For a fixed set of parameters the basic reproductive number, \mathcal{R}_0 , can be calculated as $E(X'_v)$, i.e. the average number of secondary infections caused by a post-index case. That is, \mathcal{R}_0 is given by

$$\mathcal{R}_0 = \sum_{x'_i=0}^\infty x'_i \sum_{k > x'_i}^\infty q'_k \pi_{k-1}(x'_i). \quad (4)$$

4.3. Example. To illustrate, we assume that the degree distribution is Poisson with mean parameter λ and the recovery distribution is exponential with rate parameter γ . This implies

$$f(t_i) = \gamma e^{-\gamma t_i}$$

and

$$q_{d_i} = \frac{\lambda^{d_i} e^{-\lambda}}{d_i!}.$$

Therefore, by Eq. (2), the offspring distribution for an index case is given by

$$P(X_i = x_i | \lambda, \beta, \gamma, \delta) = \sum_{k \geq x_i} \frac{\lambda^k e^{-\lambda}}{k!} \int_0^\infty \binom{k}{x_i} p(t; \beta, \delta)^{x_i} (1 - p(t; \beta, \delta))^{k-x_i} \gamma e^{-\gamma t} dt,$$

and, by Eq. (3), the post-index case offspring distribution is given by

$$\begin{aligned} P(X'_i = x'_i | \lambda, \beta, \gamma, \delta) \\ = \sum_{k > x'_i} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} \int_0^\infty \binom{k-1}{x'_i} p(t; \beta, \delta)^{x'_i} (1 - p(t; \beta, \delta))^{k-1-x'_i} \gamma e^{-\gamma t} dt. \end{aligned}$$

This expression has no simple analytical form but it is not hard to approximate the integral numerically since it can be written as an expectation against the recovery distribution. Therefore, a simple Monte Carlo sample from the desired recovery distribution allows for efficient computation of this term.

Using Eq. (4), we can calculate the basic reproductive number in this Markovian case. For an arbitrary degree distribution, \mathcal{R}_0 is given by

$$\mathcal{R}_0 = \frac{\beta}{\beta + \gamma + \delta} \sum_{k=0}^\infty \frac{(k-1)kq_k}{\mu}, \quad (5)$$

which only differs in the inclusion of δ from the corresponding formula on a static CM graph [24, 14]. Here $\mu = E(D) < \infty$ by assumption. The summation in Eq. (5) represents the expected excess degree, i.e. the degree of a node which is necessarily a neighbor of a node, not counting the known edge. In the particular case here of a Poisson degree distribution, \mathcal{R}_0 is found to be (cf., e.g., [3] chapter 6)

$$\mathcal{R}_0 = \frac{\beta}{\beta + \gamma + \delta} \sum_{k=0}^\infty \frac{\lambda^{k-1} e^{-\lambda}}{k!} k(k-1) = \frac{\beta \lambda}{\beta + \gamma + \delta}. \quad (6)$$

4.4. Likelihood and estimation. In practice, outbreak data may not arise solely from a single index case and often m separate chains of infection are tracked. Suppose the data $\{x_1, \dots, x_m\}$ corresponds to the number of secondary infections for each of m independent index cases and the data $\{x'_1, \dots, x'_{m'}\}$ corresponds to secondary infections caused by each of the m' post-index cases.

Let $\Theta = (\beta, \gamma, \delta, \lambda)$ denote the vector of parameters where γ and λ represent the parameters of the recovery time and degree distributions, respectively. The offspring distributions given in Eqs. (2) and (3) allow for explicit formulation of the likelihood for Θ which is given by

$$L(\Theta | x_1, \dots, x_m, x'_1, \dots, x'_{m'}) = \prod_{j=1}^m \sum_{k \geq x_j} q_k \pi_k(x_j) \times \prod_{c=1}^{m'} \sum_{l > x'_c} q'_l \pi_{l-1}(x'_c). \quad (7)$$

With the specification of the likelihood, maximum likelihood estimators (MLEs) for the rate parameters can be found by numerical optimization. Given the MLE $\hat{\theta} = (\hat{\beta}, \hat{\gamma}, \hat{\delta}, \hat{\lambda})$, the corresponding estimator for the basic reproductive number can be calculated by application of the continuous mapping theorem to the expression for \mathcal{R}_0 given in Eq. (4). For example, under the assumptions of our example in Section 4.3, the estimator following from Eq. (6) would be

$$\hat{\mathcal{R}}_0 = \frac{\hat{\beta} \hat{\lambda}}{\hat{\beta} + \hat{\gamma} + \hat{\delta}}. \quad (8)$$

Denote the vector of current parameters as $\Theta = (\beta, \gamma, \delta, \lambda)$. If we denote the parameter prior distribution $\phi(\Theta)$ then the likelihood function above may be also used to compute the Metropolis-Hastings acceptance probability in the Markov chain Monte Carlo (MCMC) sampler. Let \mathbf{x} denote the vector of data counts and τ be the transition kernel. The MCMC algorithm for obtaining the posterior distribution of Θ is then as follows

1. Initiate $\Theta_{curr} = \Theta_0$.
2. Obtain proposal Θ_{prop} from $\tau(\Theta|\Theta_{cur})$.
3. Accept (or not) this proposal with Metropolis-Hastings probability given by

$$\rho(\mathbf{x}, \Theta_{cur}, \Theta_{prop}) = \min \left(1, \frac{L(\Theta_{prop}|\mathbf{x})\phi(\Theta_{prop})\tau(\Theta_{prop}|\Theta_{cur})}{L(\Theta_{cur}|\mathbf{x})\phi(\Theta_{cur})\tau(\Theta_{cur}|\Theta_{prop})} \right). \quad (9)$$

4. Return to 2.
5. Repeat until convergence.

In this way, after sampler convergence, we obtain an approximate sample from the posterior distribution of Θ and may subsequently compute its approximate $1 - \alpha$ credibility region, given the observed data \mathbf{x} . In particular, the credibility interval of the parameter \mathcal{R}_0 given by Eq. (4) may be determined.

4.5. Generalizations. Note that the likelihood formula (7) is valid in a non-Markovian setting such as an arbitrary recovery time distribution and could further be extended to the scenario where the transmission rate varies with time since infection. Note that in the latter case the formula for \mathcal{R}_0 would differ from Eq. (4) due to a form for p_{t_i} that differs from Eq. (1) but would remain calculable as $E(X'_v)$.

If additional data were available, such as the recovery time or number of contacts of each individual, it could be explicitly incorporated into the likelihood function (7) through the joint distribution of recovery times and degrees.

5. Analysis of the DRC dataset. To illustrate our method we perform the Bayesian posterior estimation of the model parameters for the 2014 Ebola outbreak in the DRC based on the data described in Section 2. The specific model considered is as given in Section 3, where transmission occurs according to an exponential distribution with rate parameter β , and the degree distribution is Poisson with parameter λ . Recovery time is assumed to follow a gamma distribution $\Gamma(\alpha, \beta)$. We note that, while incubation periods for Ebola range from two to 21 days [9], our method does not require consideration of latent exposure since it does not depend on infection timing.

We perform estimation via the MCMC scheme given in Section 4.4. Prior distributions were set to be minimally informative Gaussian distributions and hyper-parameters were selected based on previous estimates [22, 30]. The prior distribution for λ was taken to be $N(16, 14)$ and for δ was taken to be $N(.01, .05)$. To improve mixing of the MCMC scheme, β was estimated on log-scale under an assumed $N(\log(.02), 4)$ prior distribution. Lastly, the gamma distribution for the infectious period was chosen to have prior mode (i.e., $(\alpha - 1)/\beta$) distributed as $N(11, 6)$ and prior standard deviation $(\sqrt{\alpha}/\beta)$ as $N(6, 4)$. A Gaussian transition kernel τ was used. Central 95% credibility intervals were calculated for the parameters of interest as well as for \mathcal{R}_0 . Results are summarized in Figure 2 including histograms of posterior samples.

The 95% credibility interval for the basic reproductive number \mathcal{R}_0 is found to be (.589, 1.15) with the \mathcal{R}_0 posterior mean of .842. The latter value numerically

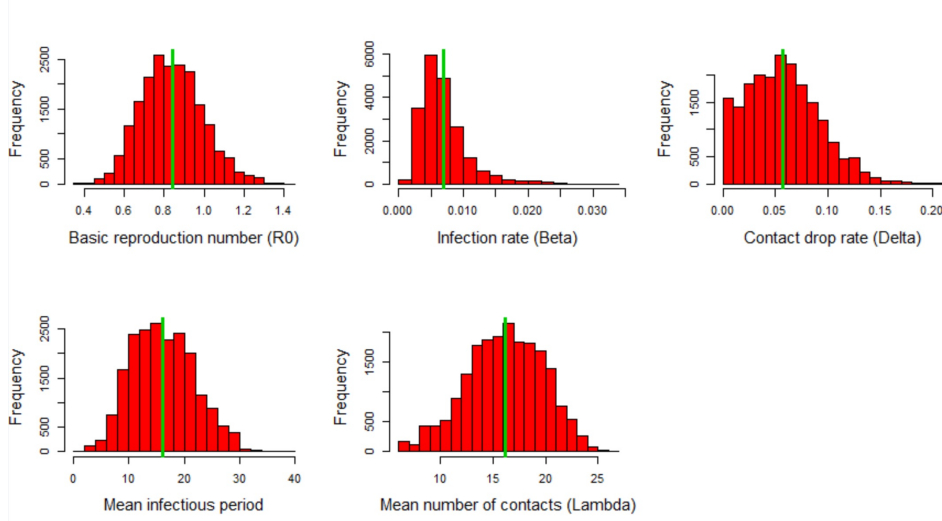


FIGURE 2. Estimated posterior densities for the parameters of interest for the non-Markovian MCMC sampler. Green line denotes posterior mean.

agrees closely with the moment-based estimate given in Maganga et al. [22] and is also consistent with the relatively small total number of confirmed infections observed during the epidemic. We further note that our interval estimate compares favorably to the 95% \mathcal{R}_0 confidence interval of $(-0.38, 2.06)$ reported by Maganga et al. [22] indicating that for the DRC data the fully parametric model produces a more precise (shorter) interval. The infection rate is found to have a posterior mean of .0069 and a corresponding 95% credibility interval of (.00232, .0167). The contact drop rate is found to have a posterior mean of .0573 and corresponding credible interval of (.00340, .128). These quantities were not estimated directly by the authors in [22] so comparison here is not possible. The estimated mean for the infectious period is found to be 16.18 days with corresponding credibility interval of (6.65 days, 27.9 days). Maganga et al. did not explicitly estimate infectious period but instead gave an estimate for time from symptom onset to death with mean 11.3 days [22]. Based on a historic 1995 outbreak of Ebola in the DRC, Legrand et al. estimated the time from onset to end of infectiousness for survivors to be 10 days and time from onset to death to be 9.6 days [21]. While our estimate of infectious period is somewhat longer than these, we note that our interpretation of the quantity is the total length of time that an individual could cause infection. This could include several days after death in which the body is being prepared for and undergoing funeral rituals. Lastly, the mean number of contacts is estimated to be 16.2 with corresponding credibility interval of (8.59, 22.8), which is again in close pointwise agreement with the estimate inferred from the number of contacts traced by Maganga et al. [22]. Overall, the point estimate for \mathcal{R}_0 and infection and recovery rates are seen to be consistent with a small outbreak behavior observed in the DRC dataset and to agree well with the numerical values reported earlier. However, based on the same data, our parametric model is also seen to yield more precise interval estimates.

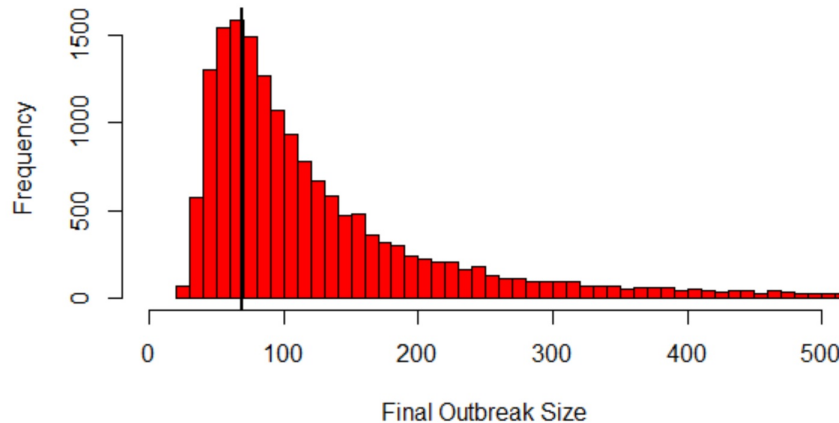


FIGURE 3. Final outbreak size distribution based on 20,000 simulations of the branching processes from the posterior parameter distribution. The actual outbreak size based on the DRC dataset (black line) is shown for comparison.

The final outbreak size of simulated branching processes from the posterior parameter distribution is used as a model diagnostic for fit to the empirical data. Conditioning on the number of infections caused by the index case as the number of independent branches, the posterior parameter samples and corresponding post-index case offspring distributions were used to simulate branching process realizations. The final outbreak sizes were calculated and the distribution is presented in Figure 3. Reasonable agreement is observed between this distribution and the empirical outbreak size of 69 cases [22].

6. Discussion. We presented here a Bayesian parameter estimation method for a class of stochastic epidemic models on configuration model graphs. The method is based on applying the branching process approximation, and is applicable when the number of infections is small in relation to the size of the population under study. In particular, this includes the case when the total outbreak size is small or when we are at the onset of a large outbreak. The method are flexible, for example allowing for arbitrary degree and recovery distributions, and in principle requires only a knowledge of the distribution of secondary cases, although additional data can be incorporated into the inference procedure, as the likelihood function under branching approximation remains straightforward to evaluate under a wide range of data collection schemes.

We illustrated our approach with the analysis of data from the 2014 DRC Ebola outbreak which was originally described and analyzed in Maganga et al. [22]. Our method, under only weakly informative prior distributions, is seen to produce a considerably tighter credibility interval for \mathcal{R}_0 than the moment-based confidence estimate reported in [22]. This demonstrates the utility of the branching process approximation for small epidemics in obtaining more precise estimates of \mathcal{R}_0 , which

is essential in assessing potential risk of a large outbreak and in determining the level of control efforts (e.g. vaccination or quarantine) needed to mitigate an outbreak. The final size comparison indicates that the observed data is within the range of model predictions, although the direct comparison of observed and model predicted offspring distributions indicates some disagreement in the observed frequency of zeros and ones. The small sample size prohibits definitive conclusions but this may indicate the need to incorporate more complex network dynamics (e.g. distinguishing between multiple types of infectious contacts).

As the statistical estimation methods appear essential to inform public health interventions, we hope that our work here will help in establishing a broader inference framework for epidemic parameters, based on the type of data usually collected in the course of an outbreak. The Bayesian approach is particularly attractive in this context, as it naturally incorporates any prior or historical information. However, the current approach only addresses the inference problem at the epidemic onset and, in particular, is not appropriate when the number of infected individuals comprises a significant portion of the population. We plan to address estimation for such large outbreaks, possibly also incorporating more complex network dynamics, in our future work.

Acknowledgments. We thank the Mathematical Biosciences Institute at OSU for its assistance in providing us with space and the necessary computational resources.

REFERENCES

- [1] L. J. Allen, An introduction to stochastic epidemic models, in *Mathematical Epidemiology*, Springer, 2008, 81–130.
- [2] M. Altmann, The deterministic limit of infectious disease models with dynamic partners, *Mathematical Biosciences*, **150** (1998), 153–175.
- [3] H. Andersson and T. Britton, *Stochastic epidemic models and their statistical analysis*, vol. 4, Springer New York, 2000.
- [4] A. D. Barbour and G. Reinert, Approximating the epidemic curve, *Electronic Journal of Probability*, **18** (2013), 1–30.
- [5] M. Boguñá, C. Castellano and R. Pastor-Satorras, Nature of the epidemic threshold for the susceptible-infected-susceptible dynamics in networks, *Physical Review Letters*, **111** (2013), 068701.
- [6] T. Bohman and M. Piccollelli, SIR epidemics on random graphs with a fixed degree sequence, *Random Structures & Algorithms*, **41** (2012), 179–214.
- [7] T. Britton and P. D. O’Neill, Bayesian inference for stochastic epidemics in populations with random social structure, *Scandinavian Journal of Statistics*, **29** (2002), 375–390.
- [8] B. Choi and G. A. Rempala, Inference for discretely observed stochastic kinetic networks with applications to epidemic modeling, *Biostatistics*, **13** (2012), 153–65.
- [9] G. Chowell and H. Nishiura, Transmission dynamics and control of Ebola virus disease (EVD): a review, *BMC Medicine*, **12** (2014), 196.
- [10] L. Decreusefond, J.-S. Dhersin, P. Moyal, V. C. Tran et al., Large graph limit for an SIR process in random network with heterogeneous connectivity, *The Annals of Applied Probability*, **22** (2012), 541–575.
- [11] C. Groendyke, D. Welch and D. R. Hunter, Bayesian inference for contact networks given epidemic data, *Scandinavian Journal of Statistics*, **38** (2011), 600–616.
- [12] C. Groendyke, D. Welch and D. R. Hunter, A network-based analysis of the 1861 Hagelloch measles data, *Biometrics*, **68** (2012), 755–765.
- [13] T. House and M. J. Keeling, Insights from unifying modern approximations to infections on networks, *Journal of The Royal Society Interface*, **8** (2011), 67–73.
- [14] S. Janson, M. Luczak and P. Windridge, Law of large numbers for the SIR epidemic on a random graph with given degrees, *Random Structures & Algorithms*, **45** (2014), 724–761.
- [15] J. Janssen, *Semi-Markov models: theory and applications*, Springer Science & Business Media, 2013.

- [16] M. Keeling, The implications of network structure for epidemic dynamics, *Theoretical Population Biology*, **67** (2005), 1–8.
- [17] M. J. Keeling, The effects of local spatial structure on epidemiological invasions, *Proceedings of the Royal Society of London B: Biological Sciences*, **266** (1999), 859–867.
- [18] M. Keeling, Correlation equations for endemic diseases: externally imposed and internally generated heterogeneity, *Proceedings of the Royal Society of London B: Biological Sciences*, **266** (1999), 953–960.
- [19] W. O. Kermack and A. G. McKendrick, A contribution to the mathematical theory of epidemics, in *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 115, The Royal Society, 1927, 700–721.
- [20] T. G. Kurtz, Limit theorems for sequences of jump markov processes approximating ordinary differential processes, *Journal of Applied Probability*, **8** (1971), 344–356.
- [21] J. Legrand, R. Grais, P. Boelle, A. Valleron and A. Flahault, Understanding the dynamics of Ebola epidemics, *Epidemiology and Infection*, **135** (2007), 610–621.
- [22] G. D. Maganga, J. Kapetshi, N. Berthet, B. Kebela Ilunga, F. Kabange, P. Mbala Kingebeni, V. Mondonge, J.-J. T. Muyembe, E. Bertherat, S. Briand et al., Ebola virus disease in the Democratic Republic of Congo, *New England Journal of Medicine*, **371** (2014), 2083–2091.
- [23] S. Meloni, N. Perra, A. Arenas, S. Gómez, Y. Moreno and A. Vespignani, Modeling human mobility responses to the large-scale spreading of infectious diseases, *Scientific Reports*, **1**.
- [24] J. C. Miller, A. C. Slim and E. M. Volz, Edge-based compartmental modelling for infectious disease spread, *Journal of the Royal Society Interface*, 890–906.
- [25] J. C. Miller and E. M. Volz, Incorporating disease and population structure into models of SIR disease in contact networks, *PloS One*, **8** (2013), e69162.
- [26] M. Newman, *Networks: an introduction*, Oxford University Press, 2010.
- [27] L. Pellis, F. Ball, S. Bansal, K. Eames, T. House, V. Isham and P. Trapman, Eight challenges for network epidemic models, *Epidemics*, **10** (2015), 58–62.
- [28] D. Rand, Correlation equations and pair approximations for spatial ecologies, *Advanced Ecological Theory: Principles and Applications*, Oxford Blackwell Science, 1999.
- [29] E. J. Schwartz, B. Choi and G. A. Rempala, Estimating epidemic parameters: Application to H1N1 pandemic data, *Math Biosciences* (2015) (e-pub, ahead of print).
- [30] W. E. R. Team, Ebola virus disease in west africa—the first 9 months of the epidemic and forward projections, *N Engl J Med*, **371** (2014), 1481–95.
- [31] E. Volz, SIR dynamics in random networks with heterogeneous connectivity, *Journal of Mathematical Biology*, **56** (2008), 293–310.
- [32] D. Welch, S. Bansal and D. R. Hunter, Statistical inference to advance network models in epidemiology, *Epidemics*, **3** (2011), 38–45.

E-mail address: burch.126@osu.edu

E-mail address: jacobsen.50@osu.edu

E-mail address: jtien@math.ohio-state.edu

E-mail address: rempala.3@osu.edu